

## Contingency and Its Two Indices Within Conditional Probability Analysis

John S. Watson

University of California at Berkeley

Four theoretical bases for detecting a contingency between behavior and consequent stimuli are considered: contiguity, correlation, conditional probability, and logical implication. It is argued that conditional probability analysis is statistically the most powerful of these options, in part due to its provision of two indices of contingency: a forward time probability that reinforcement follows behavior and a backward time probability that behavior precedes reinforcement. Evidence is cited that both indices appear to bear on the learning of a variety of animals, although they are unequally salient to human adults and to artificial neural networks designed to solve time-series functions. It is hypothesized that humans may acquire the capacity to detect contingency in the progressive sequence: contiguity, correlation, forward time conditional probability, backward time conditional probability, and ultimately logical implication.

*Key words:* contingency, infancy, learning, reinforcement

Which came first, contingency or reinforcement? I do not think this question is as fruitless or as impenetrable as the proverbial chicken or egg. Although this hypothesis is perhaps not testable, it seems likely that reinforcer power (and thus the possibility of reinforcement) evolved because contingencies existed between the behavior of organisms and beneficial changes in the environment that that behavior could affect. From this view, contingency preceded reinforcement in the evolution of learning. This “just so” story would not seem just so good were it reversed with the claim that reinforcers evolved prior to organisms evolving a capacity to affect that feature of the environment. Why would stimuli be given power to affect behavior if no contingency existed between those stimuli and the organism’s behavioral options? As Skinner noted in a discussion of “Why is a reinforcer reinforcing?” in *Science and Human Behavior* (1953),

The connection between reinforcement and satiation must be sought in the process of evolution. We can scarcely overlook the great biolog-

ical significance of the primary reinforcers. Food, water, and sexual contact, as well as escape from injurious conditions . . . , are obviously connected with the well-being of the organism. An individual who is readily reinforced by such events will acquire highly efficient behavior. (p. 83)

When framed in this light, it seems plausible that reinforcer power is no more or less than a part of one of Mother Nature’s contingency detection devices. Presumably giving reinforcing power to select stimuli directs adaptive change in behavior with respect to contingencies that bear upon the organism’s reproductive fitness. In general, this perspective views reinforcers as having evolved their power to provide reinforcement because this capacity to modify behavior frequency led to a greater engagement of contingencies that resulted in advantageous access to food, water, reproductive opportunity, and avoidance of life-terminating situations (Gewirtz & Pelaez-Nogueras, 1992).

So then, in this view, contingency is primal. Reinforcers work to direct the organism toward certain of the many contingencies that exist in the potential relationships of behavior to environmental effects. Presumably primary reinforcers direct the organism toward engagement of contingencies that bear directly on reproductive fitness. But

---

Correspondence concerning this article should be addressed to John S. Watson, Department of Psychology, University of California, Berkeley, California 94720-1650 (E-mail: jwatson@socrates.berkeley.edu).

how does this work, really? Reinforcer power can specify some contingencies as more important than others, but the simple fact that a consequence of behavior is capable of altering the strength of the behavior does not specify how this fact provides detection and engagement of that contingency. What is needed is some mechanism that either directly or indirectly detects the essential structure of the beneficial contingency. If reinforcer power does not entail contingency detection, then what else is needed to provide an adaptive mechanism for contingency detection?

To help frame this question, imagine that you were asked to design an efficient contingency-detecting system for some modern robotics device. The device would have some number of outputs and some number of inputs. Your detection system would need to detect when one or more of the outputs had an effect on one or more of the inputs. If this is to be a powerful, all-purpose system, then it should be capable of detecting contingency even when the contingent input occurs with some delay or despite the fact that it may be less than perfectly contingent (i.e., as evidenced by the fact that the input sometimes does not occur following the output that it generally is contingent upon or it occurs occasionally in the absence of the output that it generally is contingent upon). Presumably Mother Nature had this job before you. It was part of the task of providing a means of contingency detection when behavior affected the environment in a manner that altered an organism's reproductive fitness. Presumably she used Darwin's selection process to come up with a workable contingency-detection device or perhaps a number of different devices for different species (or different development levels of any species). Your task, however, is just to come up with at least one functional device. What are your options?

During the past 30 years, researchers working with human infants have centered their attention on what appear to

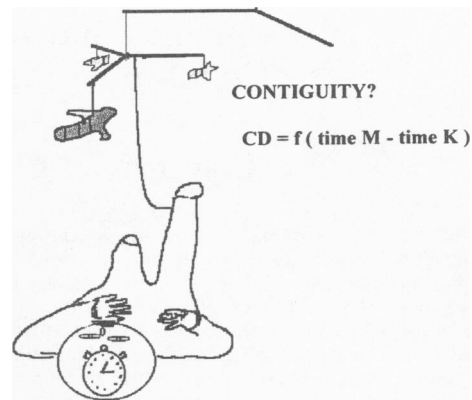


Figure 1. Depiction of contingency detection (CD) being computed by reference to temporal contiguity of mobile movement (M) and infant's kick (K).

be four different options for how contingency might be detected. I will give these the short titles of *contiguity*, *correlation*, *conditional probability*, and *logical implication*. Let us consider each in turn before I state my reasons for centering my attention on the conditional probability option.

I find it useful to illustrate these four options with a concrete situation within which they might each be applied. Imagine an infant lying in a crib. Above the infant is a mobile. One of the infant's legs has a ribbon attached to it, and that ribbon is also attached to the mobile. This situation is modeled after an experimental procedure used for many years by Rovee-Collier and her colleagues (Rovee & Rovee, 1969; Rovee-Collier, 1987). The physical arrangement provides a contingency relation between the infant's foot movement and the movement of the mobile. The question now is, on what basis might the infant detect this contingency?

### Contiguity

What appears to be the simplest method of contingency detection is assessment of temporal contiguity (see Figure 1). This was Skinner's choice. In the field of infant learning, it appears to have been Gewirtz's choice

(Gewirtz & Pelaez-Nogueras, 1992). It is an indirect detection method in the sense that it does not really compute or measure contingency per se but relies on a consequence of contingency under certain special assumptions. The special assumptions are that contingent effects of behavior usually occur with a short latency from the time of that behavior and that these same effects do not often occur noncontingently. A more familiar way of stating this perspective is to say that an effective contingency requires a reinforcer to occur within a very short time following behavior. If reinforcement is delayed too long, then it is effectively noncontingent and will have no reinforcing consequence for behavior that is truly affecting its occurrence. Another implication of this indirect method of detecting contingency is that reliance on contiguity predicts that a reinforcer will have a reinforcing effect if it occurs shortly following a behavior even by chance alone. This is the basis for so-called "superstitious" learning (Skinner, 1953). If this were possible, it would mean that a true contingency is not a necessary condition for reinforcement effects, although this logical relation appears to be the assumption of some Skinnerian theorists (Gewirtz & Pelaez-Nogueras, 1992). Whether superstitious learning actually occurs has been contested (Justice & Looney, 1990; Staddon & Simmelhag, 1970; Timberlake & Lucas, 1985), but we will not worry about that debate at this time.

The basic proposal of the contiguity perspective, then, is that the subject will show effects of the contingency as a simple effect of how frequently the consequence occurs shortly after instances of behavior. This is a reasonable basis for detecting contingency, so long as contingency between a behavior and a stimulus is expressed immediately and the stimulus is not occurring frequently by chance. A problem begins to arise when contingent effects are delayed. In the concrete example we are considering, we can surely

imagine that the mobile would move immediately after the foot movement so that the contingency would be exposed for detection by many instances of contiguity between behavior and mobile stimulation. A notable aspect of this detection system is that all it requires is a means of sensing the temporal separation of the behavior and its contingent effect. This mechanism is symbolized in Figure 1 by a stopwatch. I am not concerned with how this sensitivity is realized biologically, whether by some simple derivative of neuronal refractory processes or by some more specialized mechanism.

### *Correlation*

An alternative perspective that is less vulnerable to random contiguity is that provided by statistical correlation. In this case the subject is presumed to be sensitive to the instances in which the consequence occurs in the absence of the behavior and to instances in which the behavior occurs in the absence of the stimulus. A number of contiguous occurrences of reinforcement can be negated by a number of noncontiguous occurrences of the stimulus. More important, the use of correlation provides subjects with a potential way in which to detect delayed contingencies. If the stimulus predictably occurred 10 s after the behavior, then the amount of behavior in any period of time will be positively correlated with amount of contingent stimulation in an equal period of time beginning 10 s later. And of notable importance would be the fact that random occurrences of the stimulus would not generate a reliable correlation at any  $t + n$  delay.

Under the conditions presented in our concrete example of a baby tethered to mobile (Figure 2), we can imagine a very high correlation being derived from the temporal pattern of behavior and stimulus. Periods of time in which there were more leg kicks would be associated with greater amount of mobile movement. Given

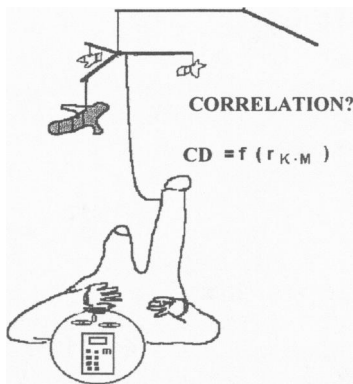


Figure 2. Depiction of contingency detection (CD) being computed by reference to correlation ( $r$ ) of amount of mobile movement (M) and amount of infant's kicking (K) over time.

that the infant could simulate the computation of this temporal correlation, the result would be expected to be positive and high.

Presumably the computation of correlation, no matter how crudely simulated, will require using more information than simply the time between behavior and stimulus. To represent this assumption of additional computational power beyond that needed by contiguity analysis, I have replaced the metaphorical stopwatch with a metaphorical hand calculator in Figure 2.

### Conditional Probability

Another option for contingency detection is the use of conditional probability analysis. This is not the most complex of the four options I will consider, but it is the one I shall elaborate on in the following sections of the paper. Its primary advantage over the preceding two options, given that it could be employed by an animal, is that it keeps track of the direction of disassociation. In the relation between any behavior and stimulus, disassociation can occur in cases of behavior not followed by the stimulus and by instances of the stimulus occurring in the absence of the behavior. Contiguity analysis is virtually blind to both forms of disassociation; it simply focuses on instances of the stimulus following the

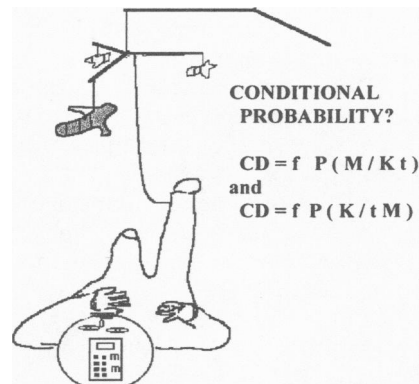


Figure 3. Depiction of contingency detection (CD) being computed by reference to the forward time or sufficiency index of conditional probability,  $p(M/Kt)$ , to be read as the probability of mobile movement (M) in a specified time span ( $t$ ) following a kick (K), and by reference to the backward time or necessity index of conditional probability,  $p(K/tM)$ , to be read as the probability of a kick (K) in a specified time span ( $t$ ) preceding a mobile movement (M).

behavior. Correlation analysis is partially sensitive to both forms of disassociation, but it mixes them together and provides a single index of association adjusted for the combined extent of disassociation.<sup>1</sup> So, for example, in the regression analysis attempting to predict a stimulus from a response, there would be just one coefficient of association such as  $r = .5$ . By contrast, conditional probability analysis provides two indices of association, each adjusted for its own direction of disassociation. Consider our baby and mobile (Figure 3). In the forward direction, the conditional probability of mobile movement *following* a leg kick might be 1.0. At the same time, in the backward direction, the conditional probability of a leg kick *preceding* a mobile movement might be only .5 if,

<sup>1</sup> I say it is only partially sensitive because sensitivity will vary with sample size of the temporal periods. For example, with sufficient sample size, intermittent reward of .5 will provide a correlation of behavior and reward across sample time periods that approximates  $r = 1.0$ . With smaller sample periods, the coefficient of correlation would begin to reflect the extent of disassociation.

say, the wind blew the mobile as often as the infant moved it by kicking. In previous papers (Watson, 1979, 1994), I have tried to demonstrate the virtual independence of these two statistical indices of contingency and the theoretical advantage of keeping track of both of them. I have also provided some evidence for the sensitivity of human infants to both indices (Watson, 1979). For the purposes of the present discussion, I will move on to complete the summary of the four focal options for contingency detection.

The details of Figure 3 include a modification of the metaphorical hand calculator. In this detection device for computing conditional probability, there is inclusion of an additional memory register over that involved in the device for the computation of correlation. This particular whimsy is meant to serve as a reminder that this detection device keeps the two directions of disassociation separate.

#### *Logical Implication*

The final option of contingency detection I will consider is that of logical implication, as expressed in symbolic logic. Bower (1989) has proposed this interesting possibility for even very young infants. Historically, a commonly held assumption has been that detection of contingency served as the basis of an inference about causality. In the present case, the direction of inference is reversed. The presence of contingency is presumed to be evaluated on the basis of the logical implications for potential causal relations of observed pairings of events. For example, if the subject observes instances of A and B as well as instances of not A and not B and then observes A and not B, Bower proposes that these observations are consistent with B implies A. They are not consistent with A implies B. Bower proposes the subject will test this hypothesis by trying to disprove it. That interesting idea about modulating behavior in order to test the hypothesis will not concern us

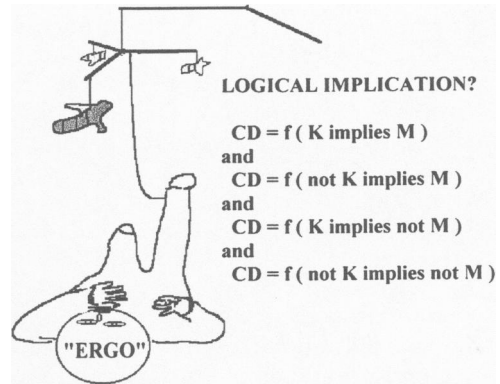


Figure 4. Depiction of contingency detection (CD) being deduced (ERGO) by reference to the logical implication of observed relations between the four causal options for combining mobile movement (M) versus no movement (not M) with kicking (K) versus no kicking (not K).

here. The point to note is simply that the subject is conceived to be using the logical relations between observed pairings of events to develop a judgment of their contingent association. In an effort to be consistent with the prior metaphors, Figure 4 replaces the memory registers with what might best be called the "sense of implication" designated by ERGO.

#### *The Advantages of Conditional Probability Analysis*

I propose that there are at least two good reasons for embracing conditional probability analysis when theorizing about the effects of variation in contingency structure and the potential evaluation of that contingency by some contingency-detection device. The first is that it embodies a plausible basis for statistical evaluation of confidence. Neither contiguity analysis nor logical implication analysis provides this advantage. Contiguity analysis simply focuses on the number of instances in which reinforcement occurs immediately following behavior without regard to the fact that this number may be that which is expected by chance. Logical implication analysis likewise fails to consider the relative probability of the categorical distinctions it makes

among logically possible pairings of events. A single instance of any category is sufficient to establish the "truth" of that pairing. The possibility of pairings occurring by chance is not considered. However, in a world in which others are capable of controlling the same outcomes as the individual in question, coincidental conjunction of the individual's behavior and stimuli produced by others seems to be an inevitable likelihood.

Correlational analysis does provide a statistical basis for evaluating the likelihood that its particular index of contingency is reliable. In the simplest manner, a particular index value is more reliable as a function of the number of pairings upon which it is based. Evaluation of reliability of a conditional probability is somewhat more complicated but is quite computable. It needs to be contrasted with the unconditioned probability of the target event. Thus, if one knows that the probability of A given B is .5, one needs to compare that value with the unconditioned probability of A, that is, the probability of A without regard to whether B has occurred. The correlation index ranges from  $-1$  to  $1$ , with difference from  $0$  being of interest, and the conditional probability ranges from  $0$  to  $1$ , with difference from unconditional probability being of interest.<sup>2</sup>

The second reason to favor conditional probability analysis is that it can (and should) make reference to the temporal characteristics of the analysis. In contiguity analysis there is an implicit concern for temporal order. If A

is said to be temporally contiguous to B, then it is conventional to interpret that statement as meaning A follows B in very little time. (Researchers understand by conventional wisdom that reinforcement does not precede the behavior it is meant to affect.) In similar fashion, logical implication analysis is concerned with the temporal order of events because it is tacitly held that in causal relations, A implies B cannot mean A causes B if A follows B (because that would imply final causation). Correlational analysis does not usually make special reference to the temporal order of the constituent events in the pairings it evaluates. However, it can and does so on occasion. In the special domain of path analysis, the temporal order of events is of major concern, but this concern is not intrinsic to the procedure of deriving a statistical coefficient of correlation. As I have argued at length elsewhere (Watson, 1979), time-based conditional probability analysis must make an explicit reference to its time base if it is to develop an appropriate evaluation of statistical reliability.

These, then, are the four basic options for contingency detection that have been the focus of researchers in the area of human infant learning. If you accepted the challenge to develop a contingency detector for a robotics device, you would be wise to choose the most powerful option you could manage to construct. Of the four we have considered, conditional probability analysis appears to be the most powerful.

An obvious question at this point is whether any of these options is employed by human infants or any other subject of learning research. There is not enough space here to review all the relevant research that bears on this question. For the purposes of this discussion I briefly note a few general observations. In research with animals, human adults, and human infants, there have been a variety of studies that have tried to assess the subject's sensitivity to statistical features of the contingen-

---

<sup>2</sup> An alternative proposal for the evaluation of conditional probability has been what is termed the difference between the conditional probability of A given B and the probability of A given the absence of B. This index also can vary between  $+1$  and  $-1$ . I have argued elsewhere (Watson, 1979) that this latter index has theoretical problems derived from its need to specify instances of nonevents. However, given its persistent popularity (Rescorla, 1967; Seligman, 1975; Shanks, 1985), I only note here that under the special assumptions needed for its application, its use carries the same implications as the index I have specified.

cy that would imply that the subject was sensitive to more than just contiguity of behavior and subsequent reward. These range from studies of dogs controlling shock (Maier, Seligman, & Solomon, 1969), to rats controlling food (Seligman, Meyer, & Testa, 1971), to infants controlling mobiles (Watson, 1971, 1977), and to college students controlling computer-generated images (Shanks, 1985, 1995). In many of these reports a conclusion is drawn that the data require an assumption that the subject is using more than just contiguity information in the process of adjusting to the contingency.

The general conclusion I draw from the array of existing studies is that all four contingency-detection options are viable and probably function in one species or another. In animals with greater computational capacity, my guess is that evolution moved to more accurate detection as provided in the move from contiguity to the statistical, or what some theorists (Shanks, 1995) call normative, evaluation. As Shanks argued, it is quite possible that these computations are approximated by simpler computational mechanisms (e.g., the delta rule in associative computation). Whether accomplished directly or by some algorithmic approximation, it is probably only in humans (perhaps only after some developmental time) that logical implication is employed in special contexts (but see Bower, 1989). Although I have no sound empirical basis for the proposal, I would not be surprised if we eventually discover that humans undergo a progressive change in their capacity to use the four contingency-detection options. I suggest that the progression is in the order I have presented these options: contiguity, followed by correlation, followed by conditional probability, followed by logical implication. I suspect that arrival of a more advanced capacity does not replace the less advanced; rather, the more advanced devices are added to the organism's overall contingency-detection repertoire. This proposal is consistent with the ob-

servation of Lewicka (1988), who has proposed that contextual factors (e.g., as in approaching vs. avoiding stimuli) may influence whether or not adult humans will show sensitivity to unconditional rates of stimuli when judging their contingent rates.

Finally, based largely on my own research (Watson, 1979, 1985), I suspect that the capacity for simulating conditional probability analysis develops in humans within the first 6 months of life, because by the age of 4 to 6 months, infants appear to show sensitivity to both forms of statistical disassociation. I would now like to propose an additional developmental transition that distinguishes the potential developmental order of capacity to use the two forms of statistical disassociation as provided by conditional probability analysis.

Before I turn to that proposal, let me summarize what I have said so far. I began by trying to point out what I take to be the primal role of contingency in the evolution of learning. Reinforcer power evolved in the evolutionary context of behavior being capable of affecting the access to and avoidance of beneficial and damaging stimuli. In addition to reinforcer power, a contingency-detection device was needed. Seemingly the simplest device was that provided with the device we call sensitivity to contiguity. This is an indirect detection device. It works insofar as contingency is generally immediate and the same stimuli are not occurring with great frequency noncontingently. The next two options we considered are what might be termed direct statistical assessments. They each evaluate the temporal dependency between behavior and stimuli, but they differ in that whereas correlation sums together both forms of disassociation, conditional probability analysis keeps these disassociations separate. The fourth option is logical implication. I proposed that although the latter was the most complex in concept, it seems vulnerable to errors of detection in a noisy, probabilistic world due to im-

precise control by multiple controllers. On the other hand, in certain contexts such as those involving rare events, logical implication may play a special role in uncovering contingencies that are difficult to apprehend with statistical analysis. I concluded by noting that it would not be surprising to me if these four options each were to be found as characteristic of some species more so than others and at some developmental stages more so than others, if not in some contexts more than others.

### *Forward Versus Backward Conditional Probability Analysis*

Let me now turn to what I see as the most interesting aspect of the option of conditional probability analysis: its potential derivation of two separate indices of contingency. Recall that by keeping the two forms of disassociation separate, this contingency-detection device generates a forward time index and a backward time index. I have elsewhere (Watson, 1979) termed these the *sufficiency* and the *necessity* indices, respectively (see Lewicka, 1988, for a similar distinction using the same terms). In the remainder of this paper, I will try to make a case for a historical and possibly causal distinction between these two indices.

Historically, it is clear that psychology has focused almost exclusively on the forward time index. When contingency was experimentally reduced, it was accomplished by reducing the probability of a stimulus reward following the organism's response. However, this manipulation only reduces the conditional probability relevant to the question "what is the probability of the reward following the behavior?" It does not reduce the probability relevant to the question "what is the probability of the behavior preceding the reward?" So long as the contingency situation does not also deliver rewards noncontingently, the backward time index will remain perfect (i.e., 1.0), regardless of how small the forward

probability becomes so long as it does not reach zero. This may be hard to believe, but consider our example of an infant with a mobile overhead. Imagine that we held the ribbon rather than attaching it to the infant's foot. Now we pull the ribbon when the infant kicks the right leg. We do this on a random intermittent schedule that provides one mobile move for two kicks on average. This results in the probability of a mobile movement being .5 immediately after a kick. But note as well that whenever a mobile movement has occurred it is always the case that it was immediately preceded by a kick. Thus, although the forward probability is .5, the backward probability is 1.

Of course the situation might be quite the reverse. Consider what would happen if we pulled the mobile ribbon each time the infant kicked, but we also pulled it randomly about an equal number of times regardless of what the infant was doing. In that case, the forward time probability would be 1 because every kick would be followed by mobile movement. But looking backward in time from the occurrence of a movement of the mobile, only half the time would there be a preceding kick. Thus, the backward probability would be only .5.

It has been only within the last 30 years that attention has been given to how schedules of reinforcement that reduce this backward probability also have adverse effects on the organism's learning (Maier et al., 1969; Seligman, 1975; Watson, 1977). The initial scientific bias toward what we are referring to as the forward probability index is also visible in the history of symbolic logic (see Copi, 1958, p. 16ff.). In that discipline, logical implication is framed in the forward direction. This is illustrated in the truth table functions that Bower (1989) applies to his model of the infant's use of logical implication. The logic is one of examining the degree to which one element is sufficient for the implied existence of the other. A implies B allows that B may exist in the absence of A but that A



cannot exist in the absence of B. Implicit is the notion that A is sufficient, but not necessary, for B. Logical implication could have been framed from the perspective of necessary relations. That would have been parallel to our concern for backward conditional probability. But it turns out that that is not the convention that evolved in symbolic logic.

The alignment of directional bias in the disciplines of psychology and logic may have been pure historical coincidence, of course. After all, given that a directional bias was to occur in each field by chance, the odds ratio would have been .5 to be aligned. On the other hand, perhaps it is not arbitrary. Perhaps there is something deeper implied.

In an effort to discover if the two indices are differentially resistant to computation, I turned to simulation as a means of approaching an answer. Mathematically, the computation of the forward time conditional probability and the backward time conditional probability are of equivalent difficulty (Watson, 1979). Psychologically, however, they seem different, as is revealed in the way it seems odd to ask about the probability of the behavior preceding the stimulus when we are considering how the behavior affected the stimulus. The question I am posing now is whether a real-time computational device would find them different.

My choice of simulation device was what is known as an artificial neural network in the recent tradition of connectionist models of brain functions and learning (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986). Because I would be simulating the analysis of events in time, I used what are called neural networks with recurrent structure (Elman, 1993). The networks have an input layer, an output layer, and a middle or so-called hidden layer. In this recurrent net, the hidden layer has a feedback connection to the input layer. As can be seen in Figure 5, the network I used has just one real

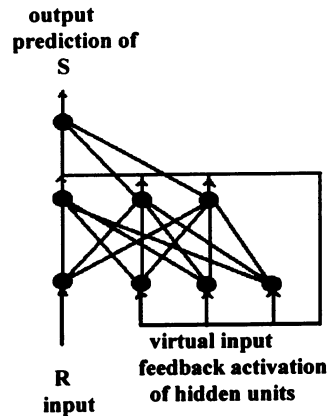


Figure 5. Schematic representation of an artificial neural network with recurrent structure. The network is composed of one output unit producing prediction of the contingent stimulus (S), three hidden-layer units that give feedback to provide three virtual input units, and one real input unit that is activated by occurrence of the effective response (R).

input unit and one output unit. In this figure the network also has three units in its hidden layer (the second level from the top in the figure). The recurrent structure that gives these networks temporal memory is shown where the three hidden units are connected to three so-called virtual input units (the rightmost three in the bottom layer of the figure).

The task for the network was this. On each trial it was informed as to whether or not a response occurred. Response occurrence activated its input node on that trial. The network either predicted the occurrence of a stimulus on that trial by activating its output unit or it didn't. A time series of 1,000 s was generated by randomly producing a response every other second on average. Then following a perfect contingency rule, a stimulus was produced after a fixed delay of 2 s for each instance of a response. So then, across time, there was no systematic relation between whether a response occurred and whether a stimulus occurred on any given second. There was also no relation to whether the stimulus occurred 1 s after a response occurred. There was a perfect relation, however,

between a response at one point in time and the occurrence of the stimulus 2 s later. Could the neural network learn to make the correct predictions regarding this contingency? The network was trained using the first 100 trials of the time series. Each epoch of training involved making a prediction on each of the 100 s. In early epochs, the network was no better than chance. But with use of the error-correction algorithm called *back propagation*, based on the error across the epoch, the net soon became a virtually perfect predictor of the stimulus on these 100 trials. This test of the net was repeated 10 times with a different randomized starting condition for its initial weight structure on each repetition. On average, it took 50 epochs of training to reach perfection on the training set. In each case, the net was also virtually perfect in its predictions across the remaining 900 s used to test its accuracy of generalization. Thus, it was clear that this recurrent net had no problem with what we have been calling forward time conditional probability. Essentially it learned to predict a stimulus (i.e., activate its output unit) given that a response occurred (i.e., its input unit had been activated) 2 s previously.

However, when the same network architecture was given the backward time problem, that of detecting whether the response systematically precedes the stimulus, it failed completely. Providing the network with information about the sequence of stimuli and asking for a prediction of response occurrence was not computable. In retrospect, the reason for this difficulty in computing the backward index of contingency is quite clear.

In the forward computation, the base of prediction is the present input and the prior inputs held in memory (i.e., the recurrent activation). The predictive error that controls learning in back propagation is generated by the difference between the network's output on a trial and the correction signal on that trial, the so-called "teacher" signal. In the present case, this is provided by

whether or not the stimulus occurs. The network soon uncovers a simple pattern relation between its memory of a response and the present stimulus.

By contrast, in the case of the backward prediction task in which the network is given the fact of a stimulus and is asked to find a relationship to responding, the network has a memory of the recent sequence of stimuli but the error is determined by the present status of whether or not a response is occurring. There is no relationship between this response and the stimulus series in memory, because the stimulus that is contingent on that response has not yet occurred. For this neural network device to work would require that it be redesigned to predict a feature of memory as opposed to a feature of the present environmental input, as is the case in a standard recurrent network. This change in design would require the device to search in memory for an appropriate teacher signal (in the connectionist terminology). Such novel design is clearly conceivable, but it is also clearly different from and more complicated than the standard recurrent network.

What might we conclude from this simulation? I think that at least one theoretically important point can be seen. The two indices of contingency that are available in conditional probability analysis, although mathematically equivalent, are not computationally equivalent when instantiated in a causal device that is by nature constrained by the unidirectional vector of time. Simple artificial neural networks can easily compute one but not the other when there is a temporal separation between behavior and contingent effects in the environment. If one accepts artificial neural networks as an approximation of brain-like devices, then it appears that backward time conditional probability analysis is a more complex act than forward time analysis. This may help to explain some existing findings with infants (Diethelm, 1991) and with adults (Lewicka, 1988) that indicate that variation in the forward

time probability appears to have greater impact than similar variation in the backward time probability. The human adult capacity to conceptualize backward time probability (as I have done in this paper) and the human infant's apparent sensitivity to manipulation of this aspect of contingency (Watson, 1979) argue for the existence of detection mechanisms at least within our species. It is enticing, though surely premature, to speculate that humans may possess a special sensitivity for backward probability relative to other animals, possibly associated with our linguistic capacity (e.g., in relation to postcedent modification). Yet, as the history of psychology and logic, as well as the few studies cited above, imply, this sensitivity appears to be a more arduous act than is that involved in the detection of forward time probability.

### General Conclusions

In this paper I have tried to explore some issues that are relevant to the concept of contingency and, in particular, to some issues regarding organismic adaptation to contingency. I have made evolutionary and ontogenetic speculations. The evolutionary speculation, probably not testable (other than in evolutionary simulation, e.g., Nolfi & Parisi, 1995), is that primary reinforcer power evolved after and as a consequence of the evolution of contingency (i.e., the evolution of effective behavior). The ontogenetic speculation, probably quite testable, is that contingency detection has at least four basic computational options, and these may well emerge in humans in a five-stage progressive sequence of apparently increasing complexity regarding what we have termed the analyses of contiguity, correlation, forward time conditional probability, backward time conditional probability, and finally logical implication.

### REFERENCES

- Bower, T. G. R. (1989). *The rational infant*. Chicago: Freeman.

- Copi, I. M. (1958). *Symbolic logic*. New York: MacMillan.
- Diethelm, K. (1991). *Mutter-kind interaktion*. Friburg, Switzerland: Hans Huber Press.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71–99.
- Gewirtz, J. L., & Pelaez-Nogueras, M. (1992). B. F. Skinner's legacy in human infant behavior and development. *American Psychologist*, 7, 1411–1422.
- Justice, T. C., & Looney, T. A. (1990). Another look at "superstitions" in pigeons. *Bulletin of the Psychonomics Society*, 28, 64–66.
- Lewicka, M. (1988). On objective and subjective anchoring of cognitive acts: How behavioral valence modifies reasoning schemata. In W. J. Baker, L. P. Mos, H. V. Rappard, & H. J. Stamm (Eds.), *Recent trends in theoretical psychology* (pp. 285–301). New York: Springer-Verlag.
- Maier, S. F., Seligman, M. E. P., & Solomon, R. L. (1969). Pavlovian fear conditioning and learned helplessness. In B. A. Campbell & R. M. Church (Eds.), *Punishment* (pp. 299–343). New York: Appleton-Century-Crofts.
- McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models*. Cambridge, MA: MIT Press.
- Nolfi, S., & Parisi, D. (1995). Evolving artificial neural networks that develop in time. In F. Moran, A. Moreno, J. J. Merelo, & P. Chacón (Eds.), *Advances in artificial life* (pp. 353–367). New York: Springer.
- Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review*, 74, 71–80.
- Rovee, C. K., & Rovee, D. T. (1969). Conjugate reinforcement of infant exploratory behavior. *Journal of Experimental Child Psychology*, 8, 33–39.
- Rovee-Collier, C. K. (1987). Learning and memory in infancy. In J. D. Osofsky (Ed.), *Handbook of infant development* (pp. 98–148). New York: Wiley.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.
- Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. San Francisco: Freeman.
- Seligman, M. E. P., Meyer, B., & Testa, T. (1971). *Appetitive helplessness: Non-contingent reinforcement retards instrumental learning*. Unpublished manuscript, University of Pennsylvania.
- Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology*, 37B, 1–21.
- Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology*, 48A, 257–279.

- Skinner, B. F. (1953). *Science and human behavior*. New York: MacMillan.
- Staddon, J. E. R., & Simmelhag, V. L. (1970). The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, 78, 3-43.
- Timberlake, W., & Lucas, G. A. (1985). The basis of superstitious behavior: Chance contingency, stimulus substitution, or appetitive behavior? *Journal of the Experimental Analysis of Behavior*, 44, 279-299.
- Watson, J. S. (1971). Cognitive-perceptual development in infancy: Setting for the seventies. *Merrill-Palmer Quarterly*, 17, 139-152.
- Watson, J. S. (1977). Depression and the perception of control in early childhood. In J. G. Shulterbrandt & A. Raskin (Eds.), *Depression in childhood: Diagnosis, treatment, and conceptual models* (pp. 123-133). New York: Raven.
- Watson, J. S. (1979). Perception of contingency as a determinant of social responsiveness. In E. B. Thoman (Ed.), *Origins of the infant's social responsiveness* (pp. 33-64). Hillsdale, NJ: Erlbaum.
- Watson, J. S. (1985). Contingency perception in early social development. In T. M. Field & N. A. Fox (Eds.), *Social perception in infants* (pp. 157-176). Norwood, NJ: Ablex.
- Watson, J. S. (1994). Detection of self: The perfect algorithm. In S. T. Parker, R. W. Mitchell, & M. L. Boccia (Eds.), *Self-awareness in animals and humans* (pp. 131-148). New York: Cambridge University Press.